



الجامعة المصرية اليابانية للعلوم والتكنولوجيا

**E-JUST**

Egypt - Japan University of Science and Technology

エジプト 日本科学技術大学

**CSE 434 - Machine Learning  
Final Project Report: Multi-Model  
Classification of Malignant Lymphoma  
Subtypes using Transfer Learning and XAI**

**By:**

Shahd Ammar  
Abdallah Adel  
Nadine Mostafa

**Submitted to:**

Prof. Waleed Gomaa  
Eng. Sara Helal

February 27, 2026

# 1 Introduction

## 1.1 Problem Statement

In critical medical classification tasks, the predictions produced by machine learning models cannot be accepted without scrutiny, as models may rely on false correlations or irrelevant contextual features rather than clinically meaningful patterns. For this reason, explainable artificial intelligence (XAI) techniques are essential to ensure the reliability and trustworthiness of medical image classification systems.

## 1.2 Solution Overview

This report presents the implementation of an image classification model for malignant lymphoma detection using a medical imaging dataset. Multiple neural network architectures are evaluated, and the Grad-CAM++ technique is applied to visualize the regions influencing the model’s predictions, ensuring that classification decisions are based on tumor-related features rather than surrounding artifacts.

# 2 Methodology

## 2.1 Dataset

## 2.2 Model Architectures

This section highlights the different model architectures used in this study.

### 2.2.1 Resnet50 pretrained on ImageNet

ResNet-50 is a deep convolutional neural network architecture that introduces residual connections to mitigate the vanishing gradient problem and enable the training of very deep models. In this study, a ResNet model pretrained on the ImageNet dataset was used to provide a comparison against DenseNet-based approaches as described later on in this section. ResNet-50 architecture is illustrated in figure 1

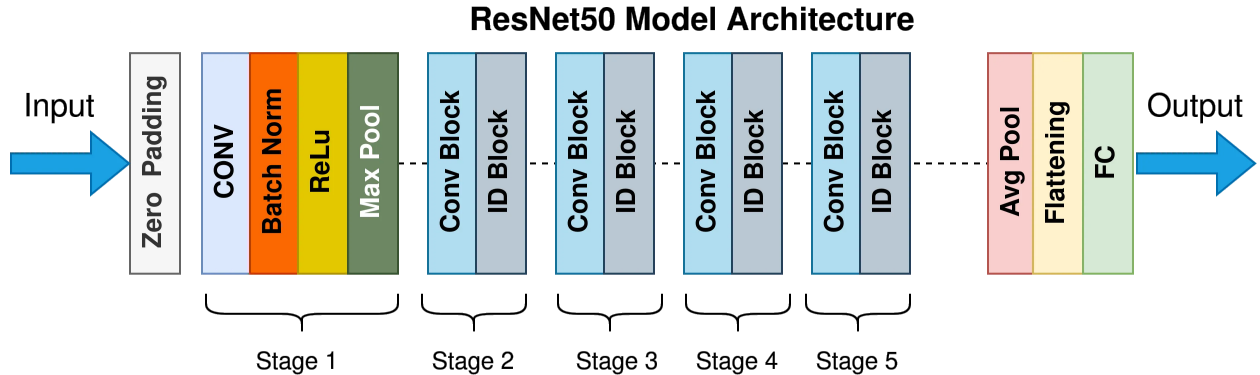


Figure 1: Resnet-50 Architecture

The task of classifying medical biopsy images for lymphoma differs significantly from the natural image classification task on which the ResNet model was originally pretrained using the ImageNet dataset.

Rather than fine-tuning only the final fully connected layer, both the final classification layer and the last convolutional block (layer4, referred to as Stage 5 in Figure 1) were fine-tuned to better adapt the model to domain-specific features present in medical images.

### 2.2.2 DenseNet-121 Architectures

To provide a robust comparison against ResNet50, two variants of the **DenseNet-121** architecture were introduced. Unlike ResNet, which uses additive identity connections, DenseNet connects each layer to every other layer in a feed-forward fashion within a "Dense Block."

The mathematical representation of a layer  $l$  in a DenseBlock is given by:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (1)$$

where  $[x_0, x_1, \dots, x_{l-1}]$  represents the concatenation of all preceding feature maps. This architecture was selected because its feature-reuse mechanism is highly effective at preserving low-level cellular textures, which are often lost in deeper residual networks.

The intuition behind the DenseBlock can be shown in figure 2

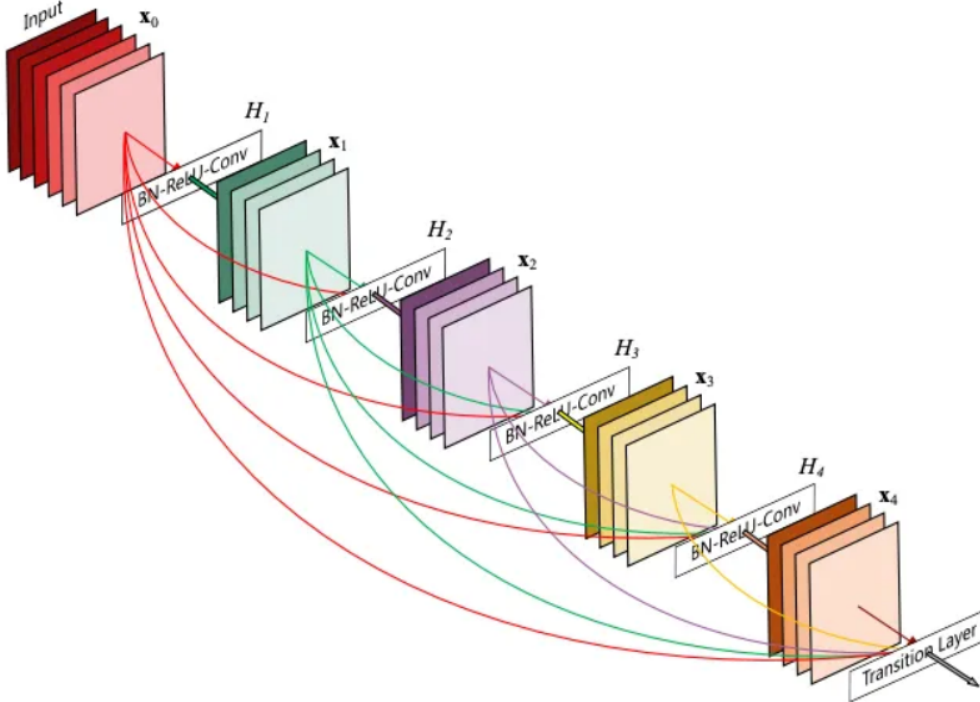


Figure 2: DenseBlock Architecture

DenseNet architecture is based on a series of dense blocks, each containing multiple convolutional layers. Each dense block takes the output of the previous block as input, as well as the outputs of all the previous blocks. This creates a dense connectivity pattern between all the layers of the network, allowing information to flow more efficiently through the network.

**DenseNet-121 (ImageNet Weights)** The first variant utilized the standard **IMAGENET1K\_V1** weights. This model serves as a benchmark for general-purpose transfer learning. The original 1000-class Softmax classifier was replaced with a linear layer tailored to the 3 lymphoma subtypes. To adapt the model to the histopathological domain, a "partial unfreezing" strategy was applied: the final dense block (*denseblock4*) was set to be trainable, while the earlier feature-extraction layers remained frozen. This allows the model to leverage general edge-detection filters while optimizing high-level spatial patterns for lymphoma cells.

**DenseNet-121 (Medical/KimiaNet Weights)** The second variant utilized domain-specific pre-training via **KimiaNet**. KimiaNet is a specialized version of DenseNet-121 trained on more than 240,000 image patches from the TCGA (The Cancer Genome Atlas) database which consisting of over 15 million pathology patches across 30 different tumor types. The reason for using KimiaNet over ImageNet is based on the *domain gap* between natural images and microscopy:

- **Texture Awareness:** ImageNet models are optimized for macroscopic object shapes. In contrast, KimiaNet is optimized for the stippled chromatin patterns, nuclear-to-cytoplasmic ratios, and pleomorphism found in digital pathology.
- **Stain Invariance:** Medical pre-training often results in weights that are more robust to the variations in Hematoxylin and Eosin (H&E) staining across different labs.

Similar to the ImageNet variant, this model was fine-tuned by unfreezing *denseblock4*, ensuring a fair and controlled comparison between the two pre-training strategies.

### 2.3 Grad-CAM++

Grad-CAM++ is an explainability method that highlights image regions most important for a network’s prediction. For a target class  $c$  and feature map  $A^k$ , it computes a weight  $\alpha_k^c$  that measures the map’s importance, using second and third derivatives of the class score to capture precise contributions. The class-specific localization map,  $L^c$ , indicates the spatial regions in the image that influence the prediction and is computed as:

$$L^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right), \quad \alpha_k^c = \frac{\sum_{i,j} \frac{\partial^2 y^c}{(\partial A_{ij}^k)^2}}{2 \sum_{i,j} \frac{\partial^2 y^c}{(\partial A_{ij}^k)^2} + \sum_{i,j} A_{ij}^k \frac{\partial^3 y^c}{(\partial (A_{ij}^k))^3}}$$

Intuitively,  $\alpha_k^c$  weights feature maps based on their influence and curvature, producing sharper heatmaps that highlight regions — such as tumor areas — that drive the model’s decision.

### 3 Results and Discussion

#### 3.1 Models Performance

The models were evaluated using four standard classification metrics: Accuracy, Precision, Recall, and F1-Score. The results summarized in Table 1 demonstrate a clear hierarchy in performance based on architecture and pre-training strategy.

Table 1: Final Performance Metrics on Test Set

Model	Accuracy	Precision	Recall	F1-Score
ResNet50 (ImageNet)	0.80	0.82	0.80	0.81
DenseNet121 (ImageNet)	0.82	0.83	0.82	0.82
DenseNet121 (Medical)	<b>0.84</b>	<b>0.85</b>	<b>0.84</b>	<b>0.84</b>

#### 3.2 Impact of Architecture: ResNet vs. DenseNet

The results show that DenseNet121 consistently outperformed ResNet50, even when using the same ImageNet initialization (82% vs 80% accuracy). This can be attributed to DenseNet’s *feature reuse* mechanism. In histopathology, critical diagnostic features exist at multiple scales—from fine nuclear textures (low-level features) to cellular arrangement patterns (high-level features). DenseNet’s dense connections ensure that low-level features are preserved and accessible to deeper layers, which appears more effective for lymphoma classification than ResNet’s summation-based residual learning.

##### 3.2.1 Impact of Domain-Specific Pre-training (KimiaNet)

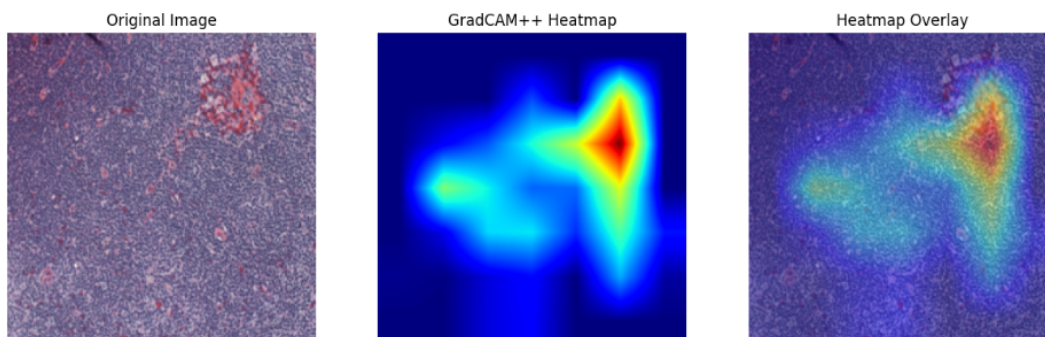
The **DenseNet Medical** model achieved the highest performance across all metrics (84% Accuracy, 0.84 F1-Score). Unlike the other models pre-trained on natural images (ImageNet), this model was initialized with KimiaNet weights, which were trained on 240,000 pathology patches.

- **Feature Relevance:** ImageNet-based models are trained to recognize edges and shapes of macroscopic objects such as cars, animals, etc. In contrast, the Medical model is already “tuned” to the stippled chromatin patterns and cytoplasmic boundaries of Hematoxylin & Eosin stained slides .
- **Convergence:** The domain-specific starting point allowed the model to reach a more optimal local minimum during fine-tuning, better distinguishing between the small morphological differences of CLL and MCL.

#### 3.3 Explainable AI (XAI) via Grad-CAM++

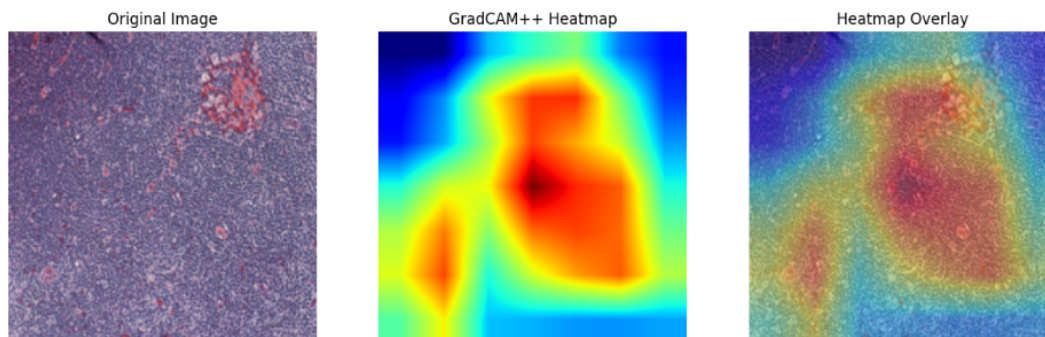
To ensure clinical relevance and validate the diagnostic integrity of the models, Grad-CAM++ was utilized to visualize the decision-making process across all three architectures. This comparative analysis shows how different pre-training strategies influence which morphological features the models prioritize.

=====  
 ResNet50 GradCAM++ Visualization  
 =====



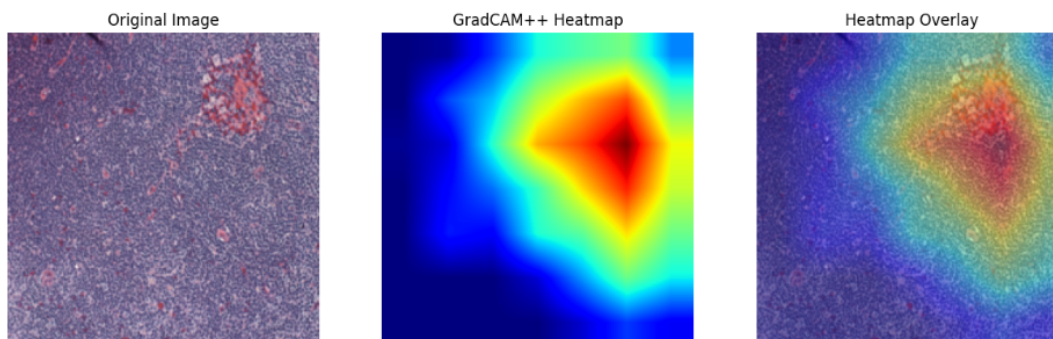
(a) ResNet50

=====  
 DenseNet ImageNet GradCAM++ Visualization  
 =====



(b) DenseNet (ImageNet)

=====  
 DenseNet Medical GradCAM++ Visualization  
 =====



(c) DenseNet (Medical/KimiaNet)

Figure 3: Comparative Grad-CAM++ visualizations. Each row illustrates how the respective model identifies diagnostic regions within the same CLL tissue sample.

### 3.3.1 Clinical Interpretability of Heatmaps

The Grad-CAM++ visualizations provided the following insights:

1. **Nuclei Focus:** The "hot" zones (red) are localized primarily over the dense lymphoid nuclei. This indicates the model is correctly identifying the hyper-cellular nature of the lymphoma rather than the connective tissue or empty spaces.
2. **Monotony Detection:** For CLL and MCL, the model's attention is spread across clusters of similar cells. This suggests the network has learned to recognize the "monotonous population" of small lymphocytes, which is a hallmark of these diseases.
3. **Artifact Immunity:** In several test cases, the model successfully ignored red blood cells and staining variations, focusing instead on the nuclear morphology of the malignant cells.

## 4 Conclusion

This project demonstrates that deep transfer learning is highly effective for lymphoma classification. While general-purpose models like ResNet50 provide high confidence, domain-specific models like DenseNet Medical offer a more better interpretation of pathological slides. The integration of XAI via Grad-CAM++ provides a transparency, proving that the models are making decisions based on relevant cellular structures, that increasing their potential for clinical decision support.

## 5 References